

Scotland's Rural College

Probabilistic forecasts: scoring rules and their decomposition and diagrammatic representation via Bregman divergences

Hughes, G; Topp, CFE

Published in:
Entropy

DOI:
[10.3390/e17085450](https://doi.org/10.3390/e17085450)

First published: 01/01/2015

Document Version
Publisher's PDF, also known as Version of record

[Link to publication](#)

Citation for pulished version (APA):

Hughes, G., & Topp, CFE. (2015). Probabilistic forecasts: scoring rules and their decomposition and diagrammatic representation via Bregman divergences. *Entropy*, 17(8), 5450 - 5471.
<https://doi.org/10.3390/e17085450>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Article

Probabilistic Forecasts: Scoring Rules and Their Decomposition and Diagrammatic Representation via Bregman Divergences

Gareth Hughes ^{†,*} and Cairistiona F.E. Topp [†]

Crop and Soil Systems, SRUC, West Mains Road, Edinburgh, EH9-3JG, UK;

E-Mail: kairsty.topp@sruc.ac.uk

[†] These authors contributed equally to this work.

* Author to whom correspondence should be addressed; E-Mail: gareth.hughes@sruc.ac.uk;
Tel.: +44-131-535-4062.

Academic Editor: Benjamin L. Ruddell

Received: 18 May 2015 / Accepted: 28 July 2015 / Published: 31 July 2015

Abstract: A scoring rule is a device for evaluation of forecasts that are given in terms of the probability of an event. In this article we will restrict our attention to binary forecasts. We may think of a scoring rule as a penalty attached to a forecast after the event has been observed. Thus a relatively small penalty will accrue if a high probability forecast that an event will occur is followed by occurrence of the event. On the other hand, a relatively large penalty will accrue if this forecast is followed by non-occurrence of the event. Meteorologists have been foremost in developing scoring rules for the evaluation of probabilistic forecasts. Here we use a published meteorological data set to illustrate diagrammatically the Brier score and the divergence score, and their statistical decompositions, as examples of Bregman divergences. In writing this article, we have in mind environmental scientists and modellers for whom meteorological factors are important drivers of biological, physical and chemical processes of interest. In this context, we briefly draw attention to the potential for probabilistic forecasting of the within-season component of nitrous oxide emissions from agricultural soils.

Keywords: scoring rule; binary forecast; Brier score; divergence score; Bregman divergence; N₂O emissions models

1. Introduction

A probabilistic forecast provides a forecast probability p that an event will subsequently occur. Probabilistic forecasts are used extensively in meteorology, so it is there that we will look for example scenarios and data. Now, qualitatively, a forecast of “rain tomorrow” with probability $p = 0.7$ means that on the basis of the forecast scheme, rain is rather more likely than not. Of course, we require definitions of “rain” and “tomorrow” in order to be able to properly interpret the forecast, but let us assume these are available. Then, given these definitions, we are able, subsequent to the forecast, to make an observation of whether or not there was rainfall in sufficient quantity to be designated “rain” during the hours designated “tomorrow”. If we view the event as binary, the outcome is either true (it rained) or false (it did not rain). Suppose it rained. From the point of view of forecast evaluation, it would be natural to give a better rating to a preceding forecast—as above—that rain was rather more likely than not ($p = 0.7$), than one that rain was less likely (*i.e.*, a smaller p). Quantitative methods for the calculation of such ratings in the context of forecast evaluation are called scoring rules [1]. This article discusses scoring rules for probabilistic forecasts. We will restrict our attention to the evaluation of forecasts for events with binary outcomes. Note that meteorologists often refer to forecast evaluation as forecast verification (e.g., [2]).

It is convenient to think of a scoring rule as a means of attaching a penalty score to a forecast; the better the forecast, the smaller the penalty (e.g., [3]). Returning to the example of a forecast of rain tomorrow with probability $p = 0.7$, the Brier score [4] is $(1 - p)^2 = 0.09$ if rain is subsequently observed and $(0 - p)^2 = 0.49$ if not. The logarithmic score (an early discussion is given in [5]) is $-\ln(p) = 0.36$ if rain is subsequently observed, and $-\ln(1 - p) = 1.20$ if not (we will use natural logarithms throughout). In practice, meteorologists are usually interested in the evaluation of a forecast scheme based on the average score for a data set comprising a sequence of forecasts and the corresponding observations. The Brier score and the logarithmic score apply different penalties; most notably, the logarithmic score attaches larger penalties than does the Brier score to forecasts for which p is close to 0 or 1 when the outcome viewed as unlikely on the basis of the forecast turns out subsequently to be the case. However, both scoring rules are “strictly proper” [6,7].

In the case of binary events, strictly proper scoring rules allow a statistical decomposition of the overall score into terms that further characterize a forecast [8]. Murphy [9] provided a statistical decomposition of the Brier score into three components, which he termed uncertainty, reliability and resolution (see also [10]). Weijs *et al.* [11,12] provided a further analysis of the logarithmic score, resulting in the divergence score and its statistical decomposition into the equivalent three components. The cited articles discuss uncertainty, reliability and resolution in detail.

Gneiting and Katzfuss [13] provide an analytical overview of probabilistic forecasting. One way of looking at the present article is as a complement to recent analytical innovations in forecast evaluation [11,12]. Using Bregman divergences, we provide a new calculation template for analysis of the Brier score and the divergence score, and new explanatory diagrams. Our objective in so doing is to provide an analysis with a straightforward diagrammatic interpretation as a basis for the evaluation of probabilistic forecasts in environmental applications where meteorological factors are important drivers of biological, physical and chemical processes of interest.

The present article is set out as follows. We introduce an example meteorological data set that is available in the public domain, and review the original analysis based on the Brier score. Following a brief discussion of the use of zero and one as probability forecasts, there is further analysis of both the Brier score and the divergence score for this data set. We then introduce our approach to the Brier score and the divergence score based on Bregman divergences, and provide examples of the calculations of the scores and their statistical decompositions. In a final discussion, we briefly mention the potential application of probabilistic forecasting to modelling of N₂O emissions from agricultural soils at the within-season time-scale.

2. Methods

2.1. Data, Terminology, Notation

In the interests of producing an analysis that allows a straightforward diagrammatic representation, we will restrict our attention here to binary outcomes. We discuss the evaluation of probability forecasts using a data set that is in the public domain. The full data set comprises 24-h and 48-h forecasts for probability of daily precipitation in the city of Tampere in south-central Finland, as made by the Finnish Meteorological Institute during 2003; together with the corresponding daily rainfall records [14]. Our analysis here is based on the 24-h rainfall forecasts. The forecasts given in [14] were made for three rainfall categories, but here, as in the original analysis, the two higher-rainfall categories were combined in order to produce a binary forecast: probability of no-rain (≤ 0.2 mm rainfall) and probability of rain (otherwise). The observations were recorded as mm precipitation but for the purpose of forecast evaluation (again as in the original analysis) the observed rainfall data were combined into the same two categories as the forecasts: observation of no-rain (≤ 0.2 mm rainfall) and observation of rain (otherwise). After excluding days for which data were missing, the full record comprised $N = 346$ probability forecasts (denoted p_t) and the corresponding observations (o_t), $t = 1, \dots, N$, with $o_t = 0$ for observation of no-rain and $o_t = 1$ for observation of rain.

The Brier score for an individual forecast is $(o_t - p_t)^2$ and the overall Brier score for a data set comprising a series of forecasts and the corresponding observations is the average of the individual scores: $BS = \frac{1}{N} \cdot \sum_{t=1}^N (o_t - p_t)^2$. This is the definition given in the original data analysis, retained for consistency. For the original data analysis the probability forecasts utilized eleven “allowed probability” forecast categories: for $k = 1, \dots, 11$; $p_k = 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$ and 1 (p_k denotes the forecast probability of rain in category k , thus the forecast probability of no-rain is its complement $1 - p_k$). The number of observations in each category is denoted n_k and the number of observations of rain in each category is denoted o_k . The average frequency of rain observations in category k is $\bar{o}_k = o_k/n_k$. Also $\sum_k n_k = N$, $\sum_k o_k = O$, and the overall average frequency of rain observations is $\bar{o} = O/N$. The components of the decomposition of the Brier score are as follows: reliability, $REL_{BS} = \frac{1}{N} \cdot \sum_k n_k \cdot (\bar{o}_k - p_k)^2$; resolution, $RES_{BS} = \frac{1}{N} \cdot \sum_k n_k \cdot (\bar{o}_k - \bar{o})^2$; uncertainty, $UNC_{BS} = \bar{o} \cdot (1 - \bar{o})$ (which is the Bernoulli variance); and then $BS = REL_{BS} - RES_{BS} + UNC_{BS}$. For the original data set, we calculate the Brier score: $BS = 0.1445$ (all calculations are shown correct to 4

d.p.). The components of the decomposition of the Brier score are: reliability, $REL_{BS} = 0.0254$; resolution, $RES_{BS} = 0.0602$; uncertainty $UNC_{BS} = 0.1793$. As required, $REL_{BS} - RES_{BS} + UNC_{BS} = BS$ and the summary of results provided along with the original data set [14] is thus reproduced.

2.2. Probability Forecasts of Zero and One

In the original data set, the probability forecasts include $p_k = 0$ (for category $k = 1$) and $p_k = 1$ (for category $k = 11$); in words, respectively, “it is certain there will be no rain tomorrow” and “it is certain there will be rain tomorrow”. Such forecasts can present problems from the point of view of evaluation. Whereas probability forecasts $0 < p_k < 1$ explicitly leave open the chance that an erroneous forecast may be made, probability forecasts $p_k = 0$ and $p_k = 1$ do not. The question that then arises is how to evaluate a forecast that was made with certainty but then proves to have been erroneous. This is not a hypothetical issue, as can be seen in the original data set. For category $k = 1$ ($p_k = 0$), we note that 1 out of the 46 forecasts made with certainty was erroneous, while for category $k = 11$ ($p_k = 1$), we note that 2 out of 13 forecasts made with certainty were erroneous [14]. If such an outcome were to occur when the logarithmic (or divergence) score was in use, an indefinitely large penalty score would apply. In routine practice our preference is to avoid the use of probability forecasts $p_k = 0$ and $p_k = 1$ (as a rule of thumb: only use a probability forecast of zero or one when there is *absolute certainty* of the outcome). There is a price to be paid for taking this point of view, which we discuss later. Notwithstanding, for further analysis in the present article, we will replace the probability forecast for category $k = 1$ by $p_k = 0.05$ (instead of zero) and the probability forecast for category $k = 11$ by $p_k = 0.95$ (instead of one) (the observations remain unchanged). A summary of the data set incorporating this adjustment (to be used exclusively from this point on) is given in Table 1.

Table 1. Summary of the data set. ^a

<i>k</i>	1	2	3	4	5	6	7	8	9	10	11
p_k	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
o_k	1	1	5	5	4	8	6	16	16	8	11
n_k	46	55	59	41	19	22	22	34	24	11	13

^a Notation: k , forecast category index; p_k , probability forecast (rain) (probability of no-rain is the complement); o_k , number of rain observations; n_k , number of observations.

2.3. The Brier Score and its Decomposition

For the adjusted data set (*i.e.*, with probability forecasts $p_k = 0.05, 0.95$ instead of 0, 1 for categories $k = 1, 11$ respectively) we recalculate the Brier score: $BS = 0.1440$. Then we recalculate the components of the decomposition of the Brier score as follows: reliability, $REL_{BS} = 0.0249$; resolution, $RES_{BS} = 0.0602$; uncertainty, $UNC_{BS} = 0.1793$. As before, $REL_{BS} - RES_{BS} + UNC_{BS} = BS$ (for full details see Appendix, Table 2).

2.4. The Divergence Score and its Decomposition

Weijs *et al.* [11,12] provide informative background on the provenance of the divergence score, and a detailed analysis of its derivation. We refer interested readers this work, and present here only enough details to illustrate a template calculation of the score and its reliability-resolution-uncertainty decomposition. The divergence score is based on the Kullback-Leibler divergence, a kind of measure of distance between two probability distributions [15,16]. For binary forecasts and the corresponding observations, all the distributions required for calculating the divergence score and its decomposition are Bernoulli, so we can write:

$$D_{KL}(x_c \| x_r) = x_c \cdot \ln \left[\frac{x_c}{x_r} \right] + (1 - x_c) \cdot \ln \left[\frac{1 - x_c}{1 - x_r} \right] \quad (1)$$

where variable x is a place-holder and, in our analysis, represents particular comparison and reference values (here, x_c and x_r , respectively) that will be replaced by a probability or a frequency, ranged between zero and one. The distribution $(x_c, 1 - x_c)$ is referred to as the comparison distribution, and the distribution $(x_r, 1 - x_r)$ is referred to as the reference distribution. Note that $D_{KL}(x_c \| x_r) \geq 0$ and that the divergence is not necessarily symmetric with respect to the arguments. For the purpose of numerical calculation, recall that $\lim_{x \rightarrow 0} [x \cdot \ln(x)] = 0$; then we take $0 \cdot \ln(0) = 0$.

The divergence score for an individual forecast is the Kullback-Leibler divergence between the observation (comparison) distribution and the forecast (reference) distribution:

$$D_{KL}(o_t \| p_t) = o_t \cdot \ln \left[\frac{o_t}{p_t} \right] + (1 - o_t) \cdot \ln \left[\frac{1 - o_t}{1 - p_t} \right].$$

For the adjusted data set we can now calculate the overall divergence score as the average of the individual scores: $DS = \frac{1}{N} \cdot \sum_{t=1}^N D_{KL}(o_t \| p_t) = 0.4471$.

The components of the decomposition of the divergence score are calculated as follows: reliability, $REL_{DS} = \frac{1}{N} \cdot \sum_k n_k \cdot D_{KL}(\bar{o}_k \| p_k) = 0.0712$; resolution, $RES_{DS} = \frac{1}{N} \cdot \sum_k n_k \cdot D_{KL}(\bar{o}_k \| \bar{o}) = 0.1683$; uncertainty (which in this case is characterized by the binary Shannon entropy [17]), $UNC_{DS} = u(\bar{o}) = -[\bar{o} \cdot \ln(\bar{o}) + (1 - \bar{o}) \cdot \ln(1 - \bar{o})] = 0.5442$. Then we have (for full details see Appendix, Table 2):

$$REL_{DS} - RES_{DS} + UNC_{DS} = DS \quad (2)$$

3. Forecast Evaluation via Bregman Divergences

Here we discuss forecast evaluation for the example data set via the Brier score and the divergence score, but using a different route through the calculations. Using Bregman divergences [18,19], our calculations lead to identical numerical results to those outlined above, in terms of the scores and their decompositions. What we gain by the analysis presented here is a set of diagrams which usefully complement those used by Weijs *et al.* [11,12] to illustrate the statistical decomposition both of the Brier score and the divergence score. This is possible because of the availability of a simple diagrammatic format for the illustration of Bregman divergences (e.g., [19,20]). So, by expressing reliability, resolution and score as Bregman divergences, we are able to illustrate these quantities

directly as distances on graphical plots. In addition, this approach enables us to write down the Brier score and the divergence score and their corresponding decompositions in a common format, thus clearly demonstrating their analytical equivalence.

Bregman divergences are properties of convex functions. In particular, the squared Euclidean distance (on which the Brier score is based) is the Bregman divergence associated with $f(x) = x^2$ and the Kullback-Leibler divergence (on which the divergence score is based) is the Bregman divergence associated with $f(x) = x \cdot \ln(x) + (1-x) \cdot \ln(1-x)$ (the negative of the binary Shannon entropy function).

Generically, a tangent to the curve $f(x)$ is drawn at x_r (the reference value). The Bregman divergence between the tangent and the curve at x_c (the comparison value) is then, for scalar arguments:

$$D_B(x_c \| x_r) = f(x_c) - f(x_r) - (x_c - x_r) \cdot f'(x_r) \quad (3)$$

in which $f'(x_r)$ is the slope of the tangent at x_r . Recall that $0 \leq x_c \leq 1$, $0 \leq x_r \leq 1$; and note that $D_B(x_c \| x_r) \geq 0$ and that the divergence is not necessarily symmetric with respect to the arguments. Where necessary for calculation purposes, we take $0 \cdot \ln(0) = 0$ as previously.

3.1. Scoring Rules as Bregman Divergences

3.1.1. Brier Score and Divergence Score Diagrams for Individual Forecast Categories

Figure 1 shows examples of scoring rules as Bregman divergences in diagrammatic form, for $p_k = 0.4$ and an observation $o \in \{0,1\}$ (see Appendix, Tables 3 and 4, category $k = 5$, for details of calculations based on Equation (3)). For individual forecasts, smaller divergences (scores) are better, and from Figure 1A (Brier score) we can see that for reference value $p_k = 0.4$ the score for comparison value $o = 0$ ($D_B = 0.16$, Table 3A, Appendix) is smaller than the score for comparison value $o = 1$ ($D_B = 0.36$, Table 3B, Appendix). From Figure 1B (divergence score) we can see that for reference value $p_k = 0.4$ the score for comparison value $o = 0$ ($D_B = 0.5108$, Table 4A, Appendix) is smaller than the score for comparison value $o = 1$ ($D_B = 0.9163$, Table 4B, Appendix). In each case this is as we require, because the forecast probability $p_k = 0.4$ is closer to $o = 0$ than to $o = 1$. That is, a forecast of $p_k = 0.4$ gets a better evaluation score if $o = 0$ is subsequently observed than if $o = 1$ is subsequently observed.

To calculate directly as Kullback-Leibler divergences the divergence scores for individual forecast categories as illustrated in Figure 1B, we have:

- for $o = 0$, $D_{KL}(0 \| p_k) = 0 \cdot \ln\left(\frac{0}{0.4}\right) + 1 \cdot \ln\left(\frac{1-0}{1-0.4}\right) = 0.5108$;
- for $o = 1$, $D_{KL}(1 \| p_k) = 1 \cdot \ln\left(\frac{1}{0.4}\right) + 0 \cdot \ln\left(\frac{1-1}{1-0.4}\right) = 0.9163$.

3.1.2. Overall Scores

For the Brier score, the Bregman divergence for each individual forecast category (as calculated via Equation (3)) is the squared Euclidean distance between o (the comparison value, where the divergence is calculated) and p_k (the reference value, where the tangent is drawn) (Appendix, Table 3). For the divergence score, the Bregman divergence for each individual forecast category (as calculated

via Equation (3)) is the Kullback-Leibler divergence between o (the comparison value, where the divergence is calculated) and p_k (the reference value, where the tangent is drawn) (Appendix, Table 4). In each case, the overall score for a forecast-observation data set is calculated as a weighted average of the individual Bregman divergences. For the Brier score, we have $BS = \frac{1}{N} \cdot \sum_k n_k \cdot D_B(o||p_k) = 49.9375/346 = 0.1440$; for the divergence score we have $DS = \frac{1}{N} \cdot \sum_k n_k \cdot D_B(o||p_k) = 154.6859/346 = 0.4471$ (for full details see Appendix, Tables 3 and 4).

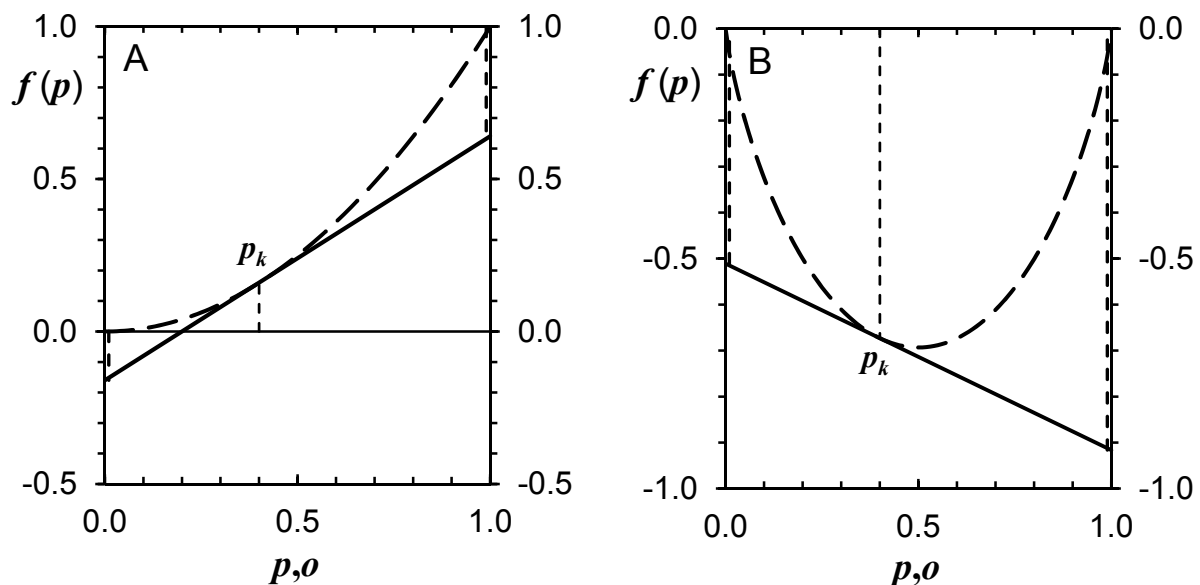


Figure 1. Scoring rules as Bregman divergences. The long-dashed curve is a convex function of p , the solid line is a tangent to the convex function at the reference value of p (p_k) indicated by a short-dashed line between the curve and the horizontal axis. The short-dashed lines between the curve and the tangent indicate the Bregman divergence at the comparison values of o (these lines coincide with sections of the vertical axes of the graphs, at comparison values $o = 0$ and $o = 1$). (A) Brier score (for calculations see Appendix, Table 3, $k = 5$). For this example, a tangent to the convex function $f(p) = p^2$ is drawn at probability forecast of rain $p_k = 0.4$. The score for this forecast depends on the subsequent observation. If no-rain is observed, the score is the Bregman divergence at $o = 0$, which is 0.16. If rain is observed, the score is the Bregman divergence at $o = 1$, which is 0.36. Bregman divergences for other forecast-observation combinations are given in the Appendix, Table 3. The overall score for a forecast-observation data set is calculated as a weighted average of the individual Bregman divergences; (B) Divergence score (for calculations see Appendix, Table 4, $k = 5$). For this example, a tangent to the convex function $f(p) = p \ln(p) + (1 - p) \ln(1 - p)$ is drawn at probability forecast of rain $p_k = 0.4$. The score for this forecast depends on the subsequent observation. If no-rain is observed, the score is the Bregman divergence at $o = 0$, which is 0.5108. If rain is observed, the score is the Bregman divergence at $o = 1$, which is 0.9163. Bregman divergences for other forecast-observation combinations are given in the Appendix, Table 4. The overall score for a forecast-observation data set is calculated as a weighted average of the individual Bregman divergences.

3.2. Reliability

3.2.1. Reliability Diagrams for Individual Forecast Categories

Figure 2 shows examples of reliability components as Bregman divergences in diagrammatic form, for reference value $p_k = 0.6$ and comparison value $\bar{o}_k = 0.2727$ (see also Appendix, Table 5, category $k = 7$, for details of calculations based on Equation (3)). From Figure 2A (for the Brier score reliability component) $D_B(\bar{o}_k \| p_k) = 0.1071$. From Figure 2B (for the divergence score reliability component) $D_B(\bar{o}_k \| p_k) = 0.2198$. The corresponding calculation for this divergence score reliability component directly as a Kullback-Leibler divergence is as follows:

$$D_{KL}(\bar{o}_k \| p_k) = 0.2727 \cdot \ln\left(\frac{0.2727}{0.6}\right) + (1 - 0.2727) \cdot \ln\left(\frac{1 - 0.2727}{1 - 0.6}\right) = 0.2198.$$

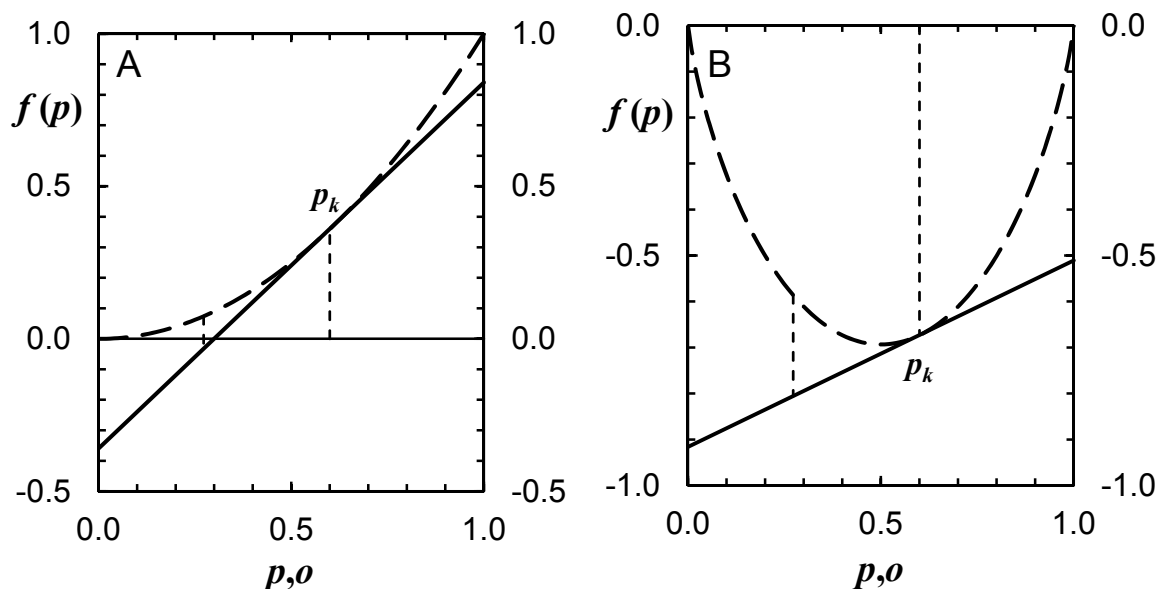


Figure 2. Reliability as a Bregman divergence. The long-dashed curve is a convex function of p , the solid line is a tangent to the convex function at the reference value of p (p_k) indicated by a short-dashed line between the curve and the horizontal axis. A second short-dashed line, between the curve and the tangent, indicates the Bregman divergence at the comparison value of o (for calculations see Appendix, Table 5). Overall reliability for a forecast-observation data set is calculated as a weighted average of individual Bregman divergences. **(A)** Brier score reliability. For this example, a tangent to the convex function $f(p) = p^2$ is drawn at probability forecast of rain $p_k = 0.6$. The reliability component depends on the corresponding \bar{o}_k , the average frequency of rain observations following such forecasts, which is 0.2727 for the example data set. The reliability component is the Bregman divergence at $\bar{o}_k = 0.2727$, which is 0.1071; **(B)** Divergence score reliability. For this example, a tangent to the convex function $f(p) = p \ln(p) + (1 - p) \ln(1 - p)$ is drawn at probability forecast of rain $p_k = 0.6$. The reliability component depends on the corresponding \bar{o}_k which is 0.2727 for the example data set. The reliability component is the Bregman divergence at $\bar{o}_k = 0.2727$, which is 0.2198.

3.2.2. Overall Reliability

For the Brier score reliability, the Bregman divergence for each individual forecast category (as calculated via Equation (3)) is the squared Euclidean distance between \bar{o}_k (the comparison value, where the divergence is calculated) and p_k (the reference value, where the tangent is drawn) (see Appendix, Table 5A). For the divergence score reliability, the Bregman divergence for each individual forecast category (as calculated via Equation (3)) is the Kullback-Leibler divergence between \bar{o}_k and p_k (see Appendix, Table 5B). In each case, the overall reliability score for a forecast-observation data set is calculated as a weighted average of the individual Bregman divergences. For the Brier score, we have $REL_{BS} = \frac{1}{N} \cdot \sum_k n_k \cdot D_B(\bar{o}_k \| p_k) = 8.6204/346 = 0.0249$; for the divergence score, we have $REL_{DS} = \frac{1}{N} \cdot \sum_k n_k \cdot D_B(\bar{o}_k \| p_k) = 24.6440/346 = 0.0712$ (for full details see Appendix, Table 5).

3.2.3. Interpreting Reliability

First, recall that reliability is defined so that smaller is better: perfect reliability corresponds to an overall reliability score equal to zero. From the formulation of the Bregman divergence $D_B(\bar{o}_k \| p_k)$, we can see that this occurs when $\bar{o}_k = p_k$ for all k categories (see Appendix, Table 5). In fact, since $D_B(\bar{o}_k \| p_k) \geq 0$, we require $\bar{o}_k = p_k$ for all k categories for an overall reliability score equal to zero. What this tells us is that for perfect reliability of our probability forecast, the average frequency of rain observations in each category must be equal to the probability forecast for that category. In practice, we typically accept (small) deviations of \bar{o}_k from p_k that contribute a small $D_B(\bar{o}_k \| p_k)$ to the overall calculation of REL_{BS} or REL_{DS} .

3.3. Resolution

3.3.1. Resolution Diagrams for Individual Forecast Categories

Figure 3 shows examples of resolution components as Bregman divergences (as calculated via Equation (3)) in diagrammatic form, for reference value $\bar{o} = 0.2341$ and comparison value $\bar{o}_k = 0.6667$ (see Appendix, Table 6, category $k = 9$). From Figure 3A (for the Brier score resolution component) $D_B(\bar{o}_k \| \bar{o}) = 0.1871$. From Figure 3B (for the divergence score resolution component) $D_B(\bar{o}_k \| \bar{o}) = 0.4204$. The corresponding calculation for this divergence score resolution component directly as a Kullback-Leibler divergence is as follows:

$$D_{KL}(\bar{o}_k \| \bar{o}) = 0.6667 \cdot \ln\left(\frac{0.6667}{0.2341}\right) + (1 - 0.6667) \cdot \ln\left(\frac{1 - 0.6667}{1 - 0.2341}\right) = 0.4204.$$

3.3.2. Overall Resolution

For the Brier score resolution, each individual Bregman divergence (as calculated via Equation (3)) is the squared Euclidean distance between \bar{o}_k (the comparison value, where the divergence is

calculated) and \bar{o} (the reference value, where the tangent is drawn) (see Appendix, Table 6A). For the divergence score resolution, each individual Bregman divergence (as calculated via Equation (3)) is the Kullback-Leibler divergence between \bar{o} and \bar{o}_k (see Appendix, Table 6B). In each case, the overall resolution score for a forecast-observation data set is calculated as a weighted average of the individual Bregman divergences. For the Brier score, we have $RES_{BS} = \frac{1}{N} \cdot \sum_k n_k \cdot D_B(\bar{o}_k \| \bar{o}) = 20.8205/346 = 0.0602$; for the divergence score we have, $RES_{DS} = \frac{1}{N} \cdot \sum_k n_k \cdot D_B(\bar{o}_k \| \bar{o}) = 58.2471/346 = 0.1683$ (for full details see Appendix, Table 6).

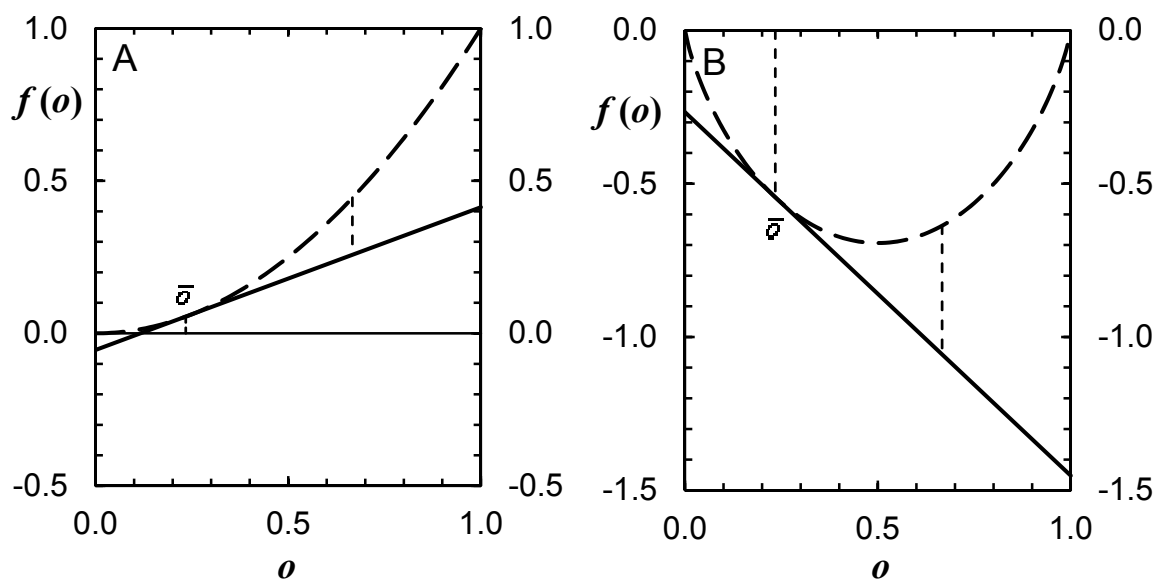


Figure 3. Resolution as a Bregman divergence. The long-dashed curve is a convex function of o , the solid line is a tangent to the convex function at the reference value of o (\bar{o}) indicated by a short-dashed line between the curve and the horizontal axis. A second short-dashed line, between the curve and the tangent, indicates the Bregman divergence at the comparison value of o (for calculations see Appendix, Table 6). Overall resolution based on a forecast-observation data set is calculated as a weighted average of the individual Bregman divergences. **(A)** Brier score resolution. For this example, a tangent to the convex function $f(o) = o^2$ is drawn at the overall average frequency of rain observations, $\bar{o} = 0.2341$. The components of resolution are calculated for each particular \bar{o}_k , the average frequency of rain observations in each category. For $k = 9$, $\bar{o}_k = 0.6667$ for the example data set. The corresponding resolution component is the Bregman divergence at $\bar{o}_k = 0.6667$, which is 0.1871; **(B)** Divergence score resolution. For this example, a tangent to the convex function $f(o) = o \cdot \ln(o) + (1 - o) \cdot \ln(1 - o)$ is drawn at the overall average frequency of rain observations, $\bar{o} = 0.2341$. The components of resolution are calculated for each particular \bar{o}_k , the average frequency of rain observations in each category. For $k = 9$, $\bar{o}_k = 0.6667$ for the example data set. The corresponding resolution component is the Bregman divergence at $\bar{o}_k = 0.6667$, which is 0.4204.

3.3.3. Interpreting Resolution

Recall that resolution is defined so that larger is better. If forecasts and observations were independent (which is least desirable), resolution would be equal to zero; if forecasts were perfect (which is most desirable), resolution would be equal to uncertainty. Note that the conditions under which resolution is equal to uncertainty also fulfil the conditions for perfect reliability, equal to zero (as above, in the context of interpreting reliability).

Resolution depends on our ability to define forecast categories for which the observed frequencies \bar{o}_k are different from the overall average frequency \bar{o} , such that the average for a forecast category provides a better prediction of the eventual outcome than the average over all forecast categories. For both the Brier score and the divergence score, if any \bar{o}_k is equal to \bar{o} , then the corresponding resolution component is equal to zero. If $\bar{o}_k = \bar{o}$ for all k , then overall resolution is equal to zero.

Consider first the scenario in which—as in the initial analysis of the original data set—probability forecasts of $p_k = 0$ and $p_k = 1$ are allowed. Further, let us suppose that all 265 observations of no-rain followed forecasts of $p_k = 0$ (in which case $\bar{o}_k = 0$) and all 81 observations of rain followed forecasts of $p_k = 1$ (so $\bar{o}_k = 1$). Recall $\bar{o} = 0.2341$. If we calculate resolution based on squared Euclidean distance, we have $RES_{BS} = \frac{1}{N} \cdot [265 \cdot (0 - \bar{o})^2 + 81 \cdot (1 - \bar{o})^2] = 62.0366/346 = 0.1793 = UNC_{BS}$. Alternatively, if we calculate resolution based on the Kullback-Leibler divergence, we have $RES_{DS} = \frac{1}{N} \cdot [265 \cdot D_{KL}(0 \| \bar{o}) + 81 \cdot D_{KL}(1 \| \bar{o})] = 188.2875/346 = 0.5442 = UNC_{DS}$. That is to say, if we were to allow probability forecast categories $p_k = 0$ and $p_k = 1$, then use them exclusively in making forecasts and do so without error, resolution would be equal to uncertainty (*i.e.*, $RES_{BS} = UNC_{BS}$ and $RES_{DS} = UNC_{DS}$).

Now consider instead the scenario in which—as in our analysis of the adjusted data set—the most extreme allowed probabilities are $p_k = 0.05$ and $p_k = 0.95$. Now, the best resolution we can achieve is if all 265 observations of no-rain followed forecasts of $p_k = 0.05$ (in which case $\bar{o}_k = 0.05$) and all 81 observations of rain followed forecasts of $p_k = 0.95$ (so $\bar{o}_k = 0.95$). If we calculate resolution based on squared Euclidean distance, we have $RES_{BS} = \frac{1}{N} \cdot [265 \cdot (0.05 - \bar{o})^2 + 81 \cdot (0.95 - \bar{o})^2] = 50.4960/346 = 0.1459$. Alternatively, if we calculate resolution based on the Kullback-Leibler divergence, we have $RES_{DS} = \frac{1}{N} \cdot [265 \cdot D_{KL}(0.05 \| \bar{o}) + 81 \cdot D_{KL}(0.95 \| \bar{o})] = 130.5177/346 = 0.3772$. Thus, the price we pay for restricting the extreme allowed probabilities to $p_k = 0.05$ and $p_k = 0.95$ is to reduce the achievable upper limit of resolution.

In the present example the notional upper limit is reduced to about 80% of uncertainty for calculations based on squared Euclidean distance, and about 70% of uncertainty for calculations based on Kullback-Leibler divergence. The difference arises because of the larger penalty score that accrues with extreme discrepancies between forecast and observation for the divergence score compared with the Brier score (as mentioned in the Introduction).

We note in passing that overall resolution, as formulated, may be characterized as a Jensen gap [21] for a convex function. Banerjee *et al.* [22] refer to this as the Bregman information. Thus generically we have $\bar{f}(x) - f(\bar{x}) \geq 0$, and in particular here, $\left[\frac{1}{N} \cdot \sum_k n_k \cdot f(\bar{o}_k) \right] - f(\bar{o}) = RES$. Then, with $f(x) = x^2$ (for the Brier score) we have $RES = \frac{1}{N} \cdot \sum_k n_k \cdot (\bar{o}_k - \bar{o})^2$, the sample variance (e.g., [3]). With $f(x) = x \cdot \ln(x) + (1-x) \cdot \ln(1-x)$ (for the divergence score) we have $RES = \frac{1}{N} \cdot \sum_k n_k \cdot D_{KL}(\bar{o}_k \| \bar{o})$, the expected mutual information (see also [11,12]).

3.4. Uncertainty

We select an uncertainty function appropriate for the analysis, depending on the chosen convex function and its associated Bregman divergence. For the Brier score, uncertainty is calculated as the value of the uncertainty function (the Bernoulli variance) at \bar{o} : $UNC_{BS} = u(\bar{o}) = \bar{o} \cdot (1 - \bar{o}) = 0.1793$ (Figure 4A). For the divergence score, uncertainty is calculated as the value of the uncertainty function (the binary Shannon entropy) at \bar{o} : $UNC_{DS} = u(\bar{o}) = -[\bar{o} \ln(\bar{o}) + (1 - \bar{o}) \ln(1 - \bar{o})] = 0.5442$ (Figure 4B). We interpret uncertainty as a quantification of our state of knowledge in the absence of a forecast, so based only on the data set from which overall average frequency of rain observations \bar{o} is calculated.

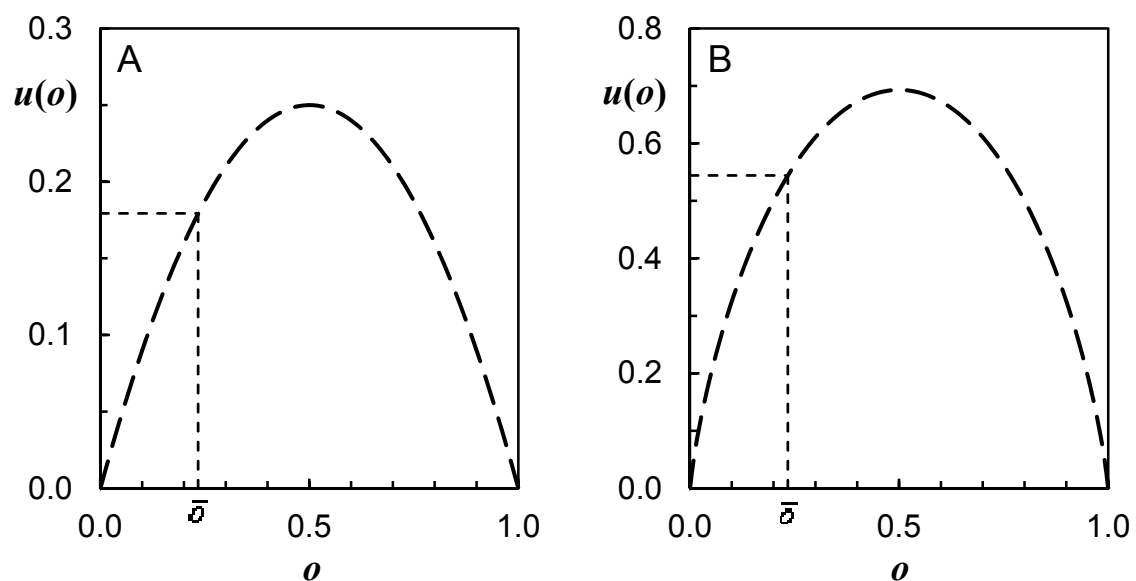


Figure 4. Uncertainty functions. The long-dashed curves are uncertainty functions, $u(o)$; the short dashed lines indicate \bar{o} ($= 0.2341$ for the example data set) and the corresponding value of $u(\bar{o})$. (A) The Bernoulli variance $u(o) = o \cdot (1 - o)$. For the example data set, $u(\bar{o}) = 0.1793$; (B) The Shannon entropy $u(o) = -(o \cdot \ln(o) + (1 - o) \cdot \ln(1 - o))$. For the example data set, $u(\bar{o}) = 0.5442$.

3.5. Overview

Theil [23] used a logarithmic scoring rule to describe the inaccuracy of predictions, but also found it convenient to write prediction errors directly in terms of the difference between the observed and forecast probabilities. This was achieved by use of a Taylor series expansion to write a logarithmic scoring rule in terms of a quadratic approximation. More recently, Benedetti [24] has attributed the lasting application of the Brier score in forecast evaluation to its being an approximation of the logarithmic score; however, an analysis leading to the Brier score as an approximation of the logarithmic score does not reveal a hierarchy in which the latter is in some way more fundamental than the former (*cf.* [25]).

For an individual probability forecast, with p_k an allowed probability and $o \in \{0,1\}$ the corresponding observation, we can calculate the scoring rule:

$$D_B(o \| p_k) = f(o) - f(p_k) - (o - p_k) \cdot f'(p_k) \quad (4)$$

(see Figure 1). Equation (4) calculates either the Brier score or the divergence score, depending on our choice of convex function on which to base the Bregman divergence. For a data set comprising a number of forecasts and corresponding observations, we calculate the overall score as $\frac{1}{N} \sum_k n_k D_B(o \| p_k)$ for either the Brier score or the divergence score. On this basis, neither scoring rule is inherently superior to the other. However, it is possible to establish further criteria against which the properties of such scoring rules may be judged [24].

The statistical decomposition of the scoring rule in Equation (4) also has a common format:

$$\left. \begin{aligned} REL_k &= D_B(\bar{o}_k \| p_k) = f(\bar{o}_k) - f(p_k) - (\bar{o}_k - p_k) \cdot f'(p_k) \\ RES_k &= D_B(\bar{o}_k \| \bar{o}) = f(\bar{o}_k) - f(\bar{o}) - (\bar{o}_k - \bar{o}) \cdot f'(\bar{o}) \\ UNC &= u(\bar{o}) \end{aligned} \right\} \quad (5)$$

(see Figures 2 and 3, respectively, for example illustrations of components of REL and RES ; and Figure 4 for an illustration of UNC , which does not vary with k). Again, it is only the choice of convex function (and corresponding choice of an appropriate uncertainty function) that distinguishes the calculation of the components of the Brier score from those of the divergence score. For a data set comprising a number of forecasts and the corresponding observations, we calculate the overall reliability and overall resolution scores, respectively, as $\frac{1}{N} \cdot \sum_k n_k \cdot D_B(\bar{o}_k \| p_k)$ and $\frac{1}{N} \cdot \sum_k n_k \cdot D_B(\bar{o}_k \| \bar{o})$.

We can compare the information-theoretic analysis of a boundary-line model by Topp *et al.* [26] with the present analysis. When, as in [26], forecast probabilities are based on retrospectively-calculated relative frequencies, reliability is equal to zero (*i.e.*, perfect reliability), uncertainty is equal to the Shannon entropy, and resolution is equal to the expected mutual information. In such a retrospective analysis, a normalized version of expected mutual information may be calculated as a measure of the proportion of uncertainty in the observations that is explained by the forecasts.

4. Discussion

Figure 5 shows a diagrammatic summary of the overall divergence score and its components (see also Equation (2)), based on calculations using the example data set. Here, uncertainty (UNC) is characterized by the binary Shannon entropy at the overall average frequency of rain observations, $\bar{o} = 0.2341$. In this context, we can think of entropy as a measure of the extent of our uncertainty before use of the forecaster. A useful intuitive interpretation of reliability (REL) can be gained from the data summary set out in Table 1. There, the probabilities p_k represent the allowed probability forecasts for rain. For a perfectly reliable forecaster, the observed frequencies of rain events, \bar{o}_k/n_k , will be equal to p_k in each category k ; then $REL = 0$. Resolution (RES) is a measure of the extent to which the forecaster accounts for uncertainty (but *not* reliability), *i.e.*, $RES \leq UNC$. As mentioned above, in the case of the divergence score, resolution is characterized by expected mutual information. Then, the divergence score (DS) characterizes the uncertainty not accounted for by the forecaster ($UNC - RES$) together with the reliability (REL), so that $DS = UNC - RES + REL$.

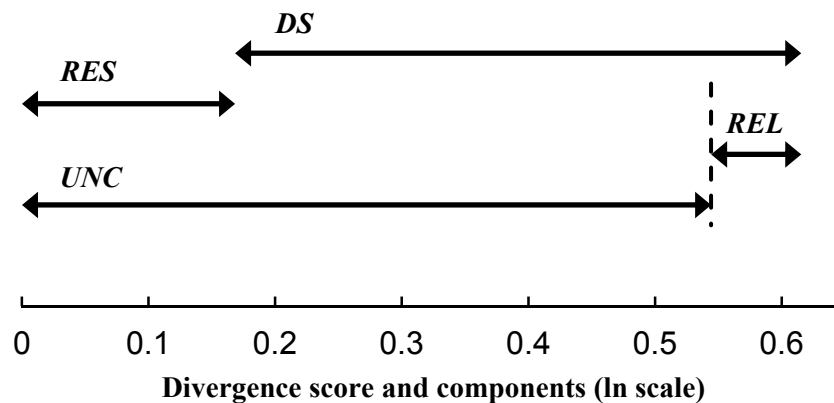


Figure 5. The overall divergence score and its components. The overall divergence score is denoted DS , with components uncertainty (UNC), reliability (REL) and resolution (RES), such that $DS = UNC - RES + REL$, with $RES \leq UNC$ as indicated by the vertical dashed line.

The evaluation of probabilistic weather forecasts is primarily of interest to meteorologists, of course; but the methodology for evaluation of probabilistic forecasts is also applicable more widely in those situations where weather factors are identified as drivers of processes contributing to risk. Weather factors are important drivers of N_2O emissions from agricultural soils, but studies of management interventions aimed at greenhouse gas mitigation have mainly been concerned with emissions inventory, and mitigation options tend to be assessed on an integrated seasonal time-scale [27,28]. An interesting example of the potential for a probabilistic approach to describing short-term N_2O flux dynamics was offered in discussion of a modelling study by Hawkins *et al.* [29], as follows: “The model depicts a realistic positive emissions response to soil moisture at the mean values of the other factors. This reflects the general understanding that N efficiency, in terms of lower N_2O emission, may be promoted by drier conditions. The WETTEST and DRIEST scenarios were simulated to investigate the magnitude of this efficiency difference. Although these scenarios are hypothetical

because in practice the wettest or driest day in a week in terms of soil moisture is not known until the end of the week, they are analogous to spreading fertiliser before or after a rainfall event.” We note here that although the wettest and driest day in a week in terms of soil moisture may only be known retrospectively, weather forecasts provide (probabilistic) advance warning of rainfall events.

Rees *et al.* [28] highlight the importance of reducing the supply of nitrogen in the context of greenhouse gas mitigation, so that management interventions with potential to increase nitrogen-use efficiency are of interest. Increasing nitrogen-use efficiency ought to represent a contribution to measures that, in relation to mitigation, reduce both greenhouse gas emissions and farm costs, constituting a “win-win” scenario [30]. The goal therefore is practical implementation of meteorological information, in the form of forecasts that could be incorporated into decision making for within-season environmental management interventions. This depends first on our ability to show that such forecasts have the required levels of reliability and resolution, using appropriate evaluation methodology as outlined here.

Acknowledgments

SRUC receives grant-in-aid from the Scottish Government.

Author Contributions

Both authors have contributed equally to this manuscript. Both authors have read and approved the final manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

Appendix

The Appendix contains the tables of results referred to in the text.

Table 2. Decomposition of the Brier score and the divergence score.^a

k	p_k	n_k	o_k	\bar{o}_k	n_k/N	$REL_{BS,k}$	$RES_{BS,k}$	$REL_{DS,k}$	$RES_{DS,k}$
1	0.05	46	1	0.0217	0.1329	0.0367	2.0745	0.4862	8.6362
2	0.1	55	1	0.0182	0.1590	0.3682	2.5642	2.9939	10.8561
3	0.2	59	5	0.0847	0.1705	0.7837	1.3162	2.9746	4.5399
4	0.3	41	5	0.1220	0.1185	1.2998	0.5157	3.6576	1.6589
5	0.4	19	4	0.2105	0.0549	0.6821	0.0106	1.5491	0.0302
6	0.5	22	8	0.3636	0.0636	0.4091	0.3691	0.8286	0.9292
7	0.6	22	6	0.2727	0.0636	2.3564	0.0328	4.8346	0.0883
8	0.7	34	16	0.4706	0.0983	1.7894	1.9014	3.8702	4.5244
9	0.8	24	16	0.6667	0.0694	0.4267	4.4907	1.1695	10.0892
10	0.9	11	8	0.7273	0.0318	0.3282	2.6754	1.3052	5.9706
11	0.95	13	11	0.8462	0.0376	0.1402	4.8699	0.9745	10.9241
Column sums ^b		346	81		1.0000	8.6204	20.8205	24.6439	58.2471

^a Notation: k , forecast category index; p_k , probability forecast (rain) (probability forecast of no-rain is the complement); n_k , number of observations; o_k , number of rain observations; \bar{o}_k , average frequency of rain observations = o_k/n_k ; n_k/N , normalized frequency of observations; $REL_{BS,k}$ (components of REL_{BS}) = $n_k \cdot (p_k - \bar{o}_k)^2$; $RES_{BS,k}$ (components of RES_{BS}) = $n_k \cdot (\bar{o}_k - \bar{o})^2$; $REL_{DS,k}$ (components of REL_{DS}) = $n_k \cdot D_{KL}(\bar{o}_k \| p_k)$; $RES_{DS,k}$ (components of RES_{DS}) = $n_k \cdot D_{KL}(\bar{o}_k \| \bar{o})$; with $\bar{o} = O/N = 0.2341$ (footnote b);

^b Column sums: $\sum_k n_k = N = 346$; $\sum_k o_k = O = 81$; $\sum_k n_k/N = 1$; $\sum_k n_k \cdot (p_k - \bar{o}_k)^2 = 8.6204$; $\sum_k n_k \cdot (\bar{o}_k - \bar{o})^2 = 20.8205$; $\sum_k n_k \cdot D_{KL}(\bar{o}_k \| p_k) = 24.6439$; $\sum_k n_k \cdot D_{KL}(\bar{o}_k \| \bar{o}) = 58.2471$.

Table 3. Brier score calculation via Bregman divergence.^aA. Observation = no-rain ($o = 0$)

k	p_k	o	n_k	$f'(p_k)$	$f(o)$	$f(p_k)$	$(o - p_k) \cdot f'(p_k)$	$D_B(0 p_k)$
1	0.05	0	45	0.1	0	0.0025	−0.0050	0.0025
2	0.1	0	54	0.2	0	0.0100	−0.0200	0.0100
3	0.2	0	54	0.4	0	0.0400	−0.0800	0.0400
4	0.3	0	36	0.6	0	0.0900	−0.1800	0.0900
5 ^b	0.4	0	15	0.8	0	0.1600	−0.3200	0.1600
6	0.5	0	14	1.0	0	0.2500	−0.5000	0.2500
7	0.6	0	16	1.2	0	0.3600	−0.7200	0.3600
8	0.7	0	18	1.4	0	0.4900	−0.9800	0.4900
9	0.8	0	8	1.6	0	0.6400	−1.2800	0.6400
10	0.9	0	3	1.8	0	0.8100	−1.6200	0.8100
11	0.95	0	2	1.9	0	0.9025	−1.8050	0.9025

B. Observation = rain ($o = 1$)

k	p_k	o	n_k	$f'(p_k)$	$f(o)$	$f(p_k)$	$(o - p_k) \cdot f'(p_k)$	$D_B(1 p_k)$
1	0.05	1	1	0.1	1	0.0025	0.0950	0.9025
2	0.1	1	1	0.2	1	0.0100	0.1800	0.8100
3	0.2	1	5	0.4	1	0.0400	0.3200	0.6400
4	0.3	1	5	0.6	1	0.0900	0.4200	0.4900
5 ^b	0.4	1	4	0.8	1	0.1600	0.4800	0.3600
6	0.5	1	8	1.0	1	0.2500	0.5000	0.2500
7	0.6	1	6	1.2	1	0.3600	0.4800	0.1600
8	0.7	1	16	1.4	1	0.4900	0.4200	0.0900
9	0.8	1	16	1.6	1	0.6400	0.3200	0.0400
10	0.9	1	8	1.8	1	0.8100	0.1800	0.0100
11	0.95	1	11	1.9	1	0.9025	0.0950	0.0025

^a Notation: k , forecast category index; p_k , probability forecast for rain (reference value, at which the tangent is calculated), probability forecast for no-rain is the complement; o , comparison value, at which the divergence is calculated; n_k , number of observations (total no-rain observations = 265, total rain observations = 81); $f'(p_k)$, slope of the tangent to $f(p)$ at p_k ; $f(o) - f(p_k) - (o - p_k) \cdot f'(p_k) = D_B(0||p_k)$ (no-rain, $o = 0$), or $f(o) - f(p_k) - (o - p_k) \cdot f'(p_k) = D_B(1||p_k)$ (rain, $o = 1$);

^b See Figure 1A.

Table 4. Divergence score calculation via Bregman divergence.^a**A.** Observation = no-rain ($o = 0$)

k	p_k	o	n_k	$f'(p_k)$	$f(o)$	$f(p_k)$	$(o - p_k) \cdot f'(p_k)$	$D_B(0 p_k)$
1	0.05	0	45	−2.9444	0	−0.1985	0.1472	0.0513
2	0.1	0	54	−2.1972	0	−0.3251	0.2197	0.1054
3	0.2	0	54	−1.3863	0	−0.5004	0.2773	0.2231
4	0.3	0	36	−0.8473	0	−0.6109	0.2542	0.3567
5 ^b	0.4	0	15	−0.4055	0	−0.6730	0.1622	0.5108
6	0.5	0	14	0.0000	0	−0.6931	0.0000	0.6931
7	0.6	0	16	0.4055	0	−0.6730	−0.2433	0.9163
8	0.7	0	18	0.8473	0	−0.6109	−0.5931	1.2040
9	0.8	0	8	1.3863	0	−0.5004	−1.1090	1.6094
10	0.9	0	3	2.1972	0	−0.3251	−1.9775	2.3026
11	0.95	0	2	2.9444	0	−0.1985	−2.7972	2.9957

B. Observation = rain ($o = 1$)

k	p_k	o	n_k	$f'(p_k)$	$f(o)$	$f(p_k)$	$(o - p_k) \cdot f'(p_k)$	$D_B(1 p_k)$
1	0.05	1	1	−2.9444	0	−0.1985	−2.7972	2.9957
2	0.1	1	1	−2.1972	0	−0.3251	−1.9775	2.3026
3	0.2	1	5	−1.3863	0	−0.5004	−1.1090	1.6094
4	0.3	1	5	−0.8473	0	−0.6109	−0.5931	1.2040
5 ^b	0.4	1	4	−0.4055	0	−0.6730	−0.2433	0.9163
6	0.5	1	8	0.0000	0	−0.6931	0.0000	0.6931
7	0.6	1	6	0.4055	0	−0.6730	0.1622	0.5108
8	0.7	1	16	0.8473	0	−0.6109	0.2542	0.3567
9	0.8	1	16	1.3863	0	−0.5004	0.2773	0.2231
10	0.9	1	8	2.1972	0	−0.3251	0.2197	0.1054
11	0.95	1	11	2.9444	0	−0.1985	0.1472	0.0513

^a Notation: k , forecast category index; p_k , probability forecast for rain (reference value, at which the tangent is calculated), probability forecast of no-rain is the complement; o , comparison value, at which the divergence is calculated; n_k , number of observations (total no-rain observations = 265, total rain observations = 81); $f'(p_k)$, slope of the tangent to $f(p)$ at p_k ; $f(o) - f(p_k) - (o - p_k) \cdot f'(p_k) = D_B(0||p_k)$ (no-rain, $o = 0$), or $f(o) - f(p_k) - (o - p_k) \cdot f'(p_k) = D_B(1||p_k)$ (rain, $o = 1$);

^b See Figure 1B.

Table 5. Reliability calculation via Bregman divergence.^a

A. Brier score

k	p_k	\bar{o}_k	n_k	$f'(p_k)$	$f(\bar{o}_k)$	$f(p_k)$	$(\bar{o}_k - p_k) \cdot f'(p_k)$	$D_B(\bar{o}_k \ p_k)$
1	0.05	0.0217	46	0.1	0.0005	0.0025	−0.0028	0.0008
2	0.1	0.0182	55	0.2	0.0003	0.0100	−0.0164	0.0067
3	0.2	0.0847	59	0.4	0.0072	0.0400	−0.0461	0.0133
4	0.3	0.1220	41	0.6	0.0149	0.0900	−0.1068	0.0317
5	0.4	0.2105	19	0.8	0.0443	0.1600	−0.1516	0.0359
6	0.5	0.3636	22	1.0	0.1322	0.2500	−0.1364	0.0186
7 ^b	0.6	0.2727	22	1.2	0.0744	0.3600	−0.3927	0.1071
8	0.7	0.4706	34	1.4	0.2215	0.4900	−0.3212	0.0526
9	0.8	0.6667	24	1.6	0.4444	0.6400	−0.2133	0.0178
10	0.9	0.7273	11	1.8	0.5289	0.8100	−0.3109	0.0298
11	0.95	0.8462	13	1.9	0.7160	0.9025	−0.1973	0.0108

B. Divergence score

k	p_k	\bar{o}_k	n_k	$f'(p_k)$	$f(\bar{o}_k)$	$f(p_k)$	$(\bar{o}_k - p_k) \cdot f'(p_k)$	$D_B(\bar{o}_k \ p_k)$
1	0.05	0.0217	46	−2.9444	−0.1047	−0.1985	0.0832	0.0106
2	0.1	0.0182	55	−2.1972	−0.0909	−0.3251	0.1798	0.0544
3	0.2	0.0847	59	−1.3863	−0.2902	−0.5004	0.1598	0.0504
4	0.3	0.1220	41	−0.8473	−0.3708	−0.6109	0.1509	0.0892
5	0.4	0.2105	19	−0.4055	−0.5147	−0.6730	0.0768	0.0815
6	0.5	0.3636	22	0.0000	−0.6555	−0.6931	0.0000	0.0377
7 ^b	0.6	0.2727	22	0.4055	−0.5860	−0.6730	−0.1327	0.2198
8	0.7	0.4706	34	0.8473	−0.6914	−0.6109	−0.1944	0.1138
9	0.8	0.6667	24	1.3863	−0.6365	−0.5004	−0.1848	0.0487
10	0.9	0.7273	11	2.1972	−0.5860	−0.3251	−0.3795	0.1187
11	0.95	0.8462	13	2.9444	−0.4293	−0.1985	−0.3058	0.0750

^a Notation: k , forecast category index; p_k , probability forecast for rain (reference value, at which the tangent is calculated), probability forecast for no-rain is the complement; \bar{o}_k , average frequency of rain observations (comparison value, at which the divergence is calculated); n_k , number of observations; $f'(p_k)$, slope of the tangent to $f(p)$ at p_k ; $f(\bar{o}_k) - f(p_k) - (\bar{o}_k - p_k) \cdot f'(p_k) = D_B(\bar{o}_k \| p_k)$;

^b See Figure 2.

Table 6. Resolution calculation via Bregman divergence.^a**A.** Brier score

k	\bar{o}	\bar{o}_k	n_k	$f'(\bar{o})$	$f(\bar{o}_k)$	$f(\bar{o})$	$(\bar{o}_k - \bar{o}) \cdot f'(\bar{o})$	$D_B(\bar{o}_k \ \bar{o})$
1	0.2341	0.0217	46	0.4682	0.0005	0.0548	−0.0994	0.0451
2	0.2341	0.0182	55	0.4682	0.0003	0.0548	−0.1011	0.0466
3	0.2341	0.0847	59	0.4682	0.0072	0.0548	−0.0699	0.0223
4	0.2341	0.1220	41	0.4682	0.0149	0.0548	−0.0525	0.0126
5	0.2341	0.2105	19	0.4682	0.0443	0.0548	−0.0110	0.0006
6	0.2341	0.3636	22	0.4682	0.1322	0.0548	0.0606	0.0168
7	0.2341	0.2727	22	0.4682	0.0744	0.0548	0.0181	0.0015
8	0.2341	0.4706	34	0.4682	0.2215	0.0548	0.1107	0.0559
9 ^b	0.2341	0.6667	24	0.4682	0.4444	0.0548	0.2025	0.1871
10	0.2341	0.7273	11	0.4682	0.5289	0.0548	0.2309	0.2432
11	0.2341	0.8462	13	0.4682	0.7160	0.0548	0.2866	0.3746

B. Divergence score

k	\bar{o}	\bar{o}_k	n_k	$f'(\bar{o})$	$f(\bar{o}_k)$	$f(\bar{o})$	$(\bar{o}_k - \bar{o}) \cdot f'(\bar{o})$	$D_B(\bar{o}_k \ \bar{o})$
1	0.2341	0.0217	46	−1.1853	−0.1047	−0.5442	0.2517	0.1877
2	0.2341	0.0182	55	−1.1853	−0.0909	−0.5442	0.2559	0.1974
3	0.2341	0.0847	59	−1.1853	−0.2902	−0.5442	0.1770	0.0769
4	0.2341	0.1220	41	−1.1853	−0.3708	−0.5442	0.1329	0.0405
5	0.2341	0.2105	19	−1.1853	−0.5147	−0.5442	0.0279	0.0016
6	0.2341	0.3636	22	−1.1853	−0.6555	−0.5442	−0.1535	0.0422
7	0.2341	0.2727	22	−1.1853	−0.5860	−0.5442	−0.0458	0.0040
8	0.2341	0.4706	34	−1.1853	−0.6914	−0.5442	−0.2803	0.1331
9 ^b	0.2341	0.6667	24	−1.1853	−0.6365	−0.5442	−0.5127	0.4204
10	0.2341	0.7273	11	−1.1853	−0.5860	−0.5442	−0.5845	0.5428
11	0.2341	0.8462	13	−1.1853	−0.4293	−0.5442	−0.7255	0.8403

^a Notation: k , forecast category index; \bar{o} , overall average frequency of rain observations (see Table 2) (reference value, at which the tangent is calculated); \bar{o}_k , average frequency of rain observations (comparison value, at which the divergence is calculated); n_k , number of observations; $f'(\bar{o})$, slope of the tangent to $f(o)$ at \bar{o} ; $f(\bar{o}_k) - f(\bar{o}) - (\bar{o}_k - \bar{o}) \cdot f'(\bar{o}) = D_B(\bar{o}_k \| \bar{o})$;

^b See Figure 3.

References

1. Lindley, D.V. *Making Decisions*, 2nd ed.; Wiley: Chichester, UK, 1985.
2. Jolliffe, I.T., Stephenson, D.B., Eds. *Forecast Verification: A Practitioner's Guide in Atmospheric Science*, 2nd ed.; Wiley: Chichester, UK, 2012.
3. Broecker, J. Probability forecasts. In *Forecast Verification: A Practitioner's Guide in Atmospheric Science*, 2nd ed.; Jolliffe, I.T., Stephenson, D.B., Eds.; Wiley: Chichester, UK, 2012; pp. 119–139.
4. Brier, G.W. Verification of forecasts expressed in terms of probability. *Mon. Weather Rev.* **1950**, *78*, 1–3.
5. Good, I.J. Rational decisions. *J. Roy. Stat. Soc. B* **1952**, *14*, 107–114.
6. DeGroot, M.W.; Fienberg, S.E. The comparison and evaluation of forecasters. *The Statistician* **1983**, *32*, 12–22.
7. Bröcker, J.; Smith, L.A. Scoring probabilistic forecasts: the importance of being proper. *Weather Forecast.* **2007**, *22*, 382–388.
8. Bröcker, J. Reliability, sufficiency, and the decomposition of proper scores. *Q. J. R. Meteorol. Soc.* **2009**, *135*, 1512–1519.
9. Murphy, A.H. A new vector partition of the probability score. *J. Appl. Meteorol.* **1973**, *12*, 595–600.
10. Wilks, D.S. *Statistical Methods in the Atmospheric Sciences*, 3rd ed.; Academic Press: Oxford, UK, 2011.
11. Weijis, S.V.; Schoups, G.; van de Giesen N. Why hydrological predictions should be evaluated using information theory. *Hydrol. Earth Syst. Sci.* **2010**, *14*, 2545–2558.
12. Weijis, S.V.; van Nooijen, R.; van de Giesen, N. Kullback-Leibler divergence as a forecast skill score with classic reliability-resolution-uncertainty decomposition. *Mon. Weather Rev.* **2010**, *138*, 3387–3399.
13. Gneiting, T.; Katzfuss, M. Probabilistic forecasting. *Annu. Rev. Stat. Appl.* **2014**, *1*, 125–151.
14. Verifying probability of precipitation—an example from Finland. <http://www.cawcr.gov.au/projects/verification/POP3/POP3.html> (accessed on 18 June 2015).
15. Kullback, S. *Information Theory and Statistics*, 2nd ed.; Dover: New York, NY, USA, 1968.
16. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; Wiley: Hoboken, NJ, USA, 2006.
17. Shannon, C.E.; Weaver, W. *The Mathematical Theory of Communication*; University of Illinois Press: Urbana, IL, USA, 1949.
18. Bregman, L.M. The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming. *USSR Comput. Math. Math. Phys.* **1967**, *7*, 200–217.
19. Gneiting, T.; Raftery, A.E. Strictly proper scoring rules, prediction, and estimation. *J. Am. Stat. Assoc.* **2007**, *102*, 359–378.
20. Adamčik, M. The information geometry of Bregman divergences and some applications in multi-expert reasoning. *Entropy* **2014**, *16*, 6338–6381.
21. Reid, M.D.; Williamson, R.C. Information, divergence and risk for binary experiments. *J. Mach. Learn. Res.* **2011**, *12*, 731–817.

22. Banerjee, A.; Merugu, S.; Dhillon, I.S.; Ghosh, J. Clustering with Bregman divergences. *J. Mach. Learn. Res.* **2005**, *6*, 1705–1749.
23. Theil, H. *Statistical Decomposition Analysis*. North-Holland Publishing Company: Amsterdam, The Netherlands, 1972.
24. Benedetti, R. Scoring rules for forecast verification. *Mon. Weather Rev.* **2010**, *138*, 203–211.
25. Tödter, J.; Ahrens, B. Generalization of the ignorance score: continuous ranked version and its decomposition. *Mon. Weather Rev.* **2012**, *140*, 2005–2017.
26. Topp, C.F.E.; Wang, W.; Cloy, J.M.; Rees, R.M.; Hughes, G. Information properties of boundary line models for N₂O emissions from agricultural soils. *Entropy* **2013**, *15*, 972–987.
27. Cardenas, L.M.; Gooday, R.; Brown, L.; Scholefield, D.; Cuttle, S.; Gilhespy, S.; Matthews, R.; Misselbrook, T.; Wang, J.; Li, C.; Hughes, G.; Lord, E. Towards an improved inventory of N₂O from agriculture: model evaluation of N₂O emission factors and N fraction leached from different sources in UK agriculture *Atmos. Environ.* **2013**, *79*, 340–348.
28. Rees, R.M.; Augustin, J.; Alberti, G.; Ball, B.C.; Boeckx, P.; Cantarel, A.; Castaldi, S.; Chirinda, N.; Chojnicki, B.; Giebel, M.; Gordon, H.; Grosz, B.; Horvath, L.; Juszczak, R.; Klemetsson, Å.K.; Klemetsson, L.; Medinets, S.; Machon, A.; Mapanda, F.; Nyamangara, J.; Olesen, J.E.; Reay, D.S.; Sanchez, L.; Sanz Cobena, A.; Smith, K.A.; Sowerby, A.; Sommer, J.M.; Soussana, J.F.; Stenberg, M.; Topp, C.F.E.; van Cleemput, O.; Vallejo, A.; Watson, C.A.; Wuta, M. Nitrous oxide emissions from European agriculture—an analysis of variability and drivers of emissions from field experiments. *Biogeosciences* **2013**, *10*, 2671–2682.
29. Hawkins, M.J.; Hyde, B.P.; Ryan, M.; Schulte, R.P.O.; Connolly, J. An empirical model and scenario analysis of nitrous oxide emissions from a fertilised and grazed grassland site in Ireland. *Nutr. Cycl. Agroecosyst.* **2007**, *79*, 93–101.
30. Moran, D.; Lucas, A.; Barnes, A. Mitigation win-win. *Nat. Clim. Change* **2013**, *3*, 611–613.